DeepGeo: Photo Localization with Deep Neural Network

Montiel Abello
mabello@andrew.cmu.edu

Nathaniel Chodosh
nchodosh@andrew.cmu.edu

Sudharshan Suresh
sudhars1@andrew.cmu.edu

Project Description and Motivation

Games such as GeoGuessr demonstrate that humans are remarkably good at estimating location from a single image. This is in spite of the fact that many (if not most) images have ambiguous locations unless they contain very specific landmarks. Humans are clearly able to use rough cues to estimate probable locations. In this project we will attempt to replicate this ability using a deep neural network.

Geo-Guesser presents the player with a Google Street View 360° panorama from anywhere in the world and asks the user to guess the GPS coordinates of the place the photo was taken. We intend to restrict the scope of this by only using images taken in the United States and classify which State they were taken in. Specifically, we will attempt to create a model which will predict a probability distribution over all States corresponding to probability that a given set of images was taken in them.

Prior Work

The challenges of geo-localization based on visual information has its foundations in [1]. Im2GPS [2] computes the closest match of an image with a large corpus of geotagged Flickr images. PlaNet [3] uses a less-restricted dataset of images to obtain a probability distribution of location over a partitioned world map. [4] scrapes data from Google Street View, but restricts the study to 2 cities. Resnet [5], based on a residual network framework, has been a go-to deep network for visual recognition.

Method and Implementation

Out dataset will be obtained via the Google Street View API, which provides a large collection of real-world imagery across the world and tools for their extraction. Using state boundaries, we will uniformly sample images from 10000 locations in each of the 50 states, for a total of 500000 examples. At each location we will extract 4 street view images pointing in the cardinal directions.

We adopt a CNN-based approach, and use a Resnet framework implemented in Tensorflow with downsampled images as input. 20% of our data will be set aside for testing. As an initial toy example, we will sample data from 5 states and quickly train a network on a subset of this data. This will allow us to assess our sampling scheme and make changes before collecting the full dataset of 2M images.

Based on these initial results, we may require a unsupervised clustering stage to categorize images into broad classes (urban, suburban and rural). Separate networks would be trained on data from these broad classes to better differentiate between the state output classes. Cross validation will be used to adjust our network architecture to achieve suitable performance. Performance of the final network will then be assessed with the testing data.

Date	**Task**
March 30	Google Street View API interface working, download small dataset
April 6	Initial results from toy example
April 11	Milestone report due
April 13	Download full dataset, run first network
April 20	Perform clustering if required and tune network for performance.
April 27	Collect results on test data. Write initial paper draft and complete poster.
April 30	Poster presentation
May 1	Finalize paper.
May 2	Final report due

References

[1] William B Thompson, Carolyn M Valiquette, BH Bennet, and Karen T Sutherland. Geometric reasoning for map-based localization. *Computer Science Technical Report UUCS-96-006, University of Utah, Salt Lake City, UT*, 1996.

[2] James Hays and Alexei A Efros. Im2gps: estimating geographic information from a single image. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[3] Tobias Weyand, Ilya Kostrikov, and James Philbin. Planet-photo geolocation with convolutional neural networks. In *European Conference on Computer Vision*, pages 37–55. Springer, 2016.

[4] Amir Roshan Zamir and Mubarak Shah. Accurate image localization based on google maps street view. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2010.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.